Project Documentation

Project title : E commerce Data analysis

Project Date: 22/11/2016

Author: Neha Solanki

Project Objective: In this e-commerce project, the basic idea is to get information of each customer for that purpose we are using table datasets. Project analysis is based on all kind of data such as structured and unstructured.

Scope : Extracted data can be used for an e commerce application to get only customer information, prediction about customer interest so that more similar choice product information can be shared and customer service can be improved.

Sample Data sets For analysis:

In this project only two dataset used to gather information for analysis purpose.

1. Customer: In customer table customer information is stored like customer id, customer first name, lastname,age of a particular customer and address . all information we want to store because we need to follow up these customer for longer time.

Customer_id	First name	last name	age	address

2. Transaction: In transaction table dataset all the transaction details are stored. Like date of transaction, user id of each customer, how much amount has paid for purchasing by each user, category, city and state of customer and payment mode of customer?

Date	uid	amount	category	product	city	state	payment

Use-Cases

Use case 1: Getting specific transactions

In this use case, we are finding all the transaction where amount is more than 160 according to customer ID, so that we come to know about the specific amount or more than that is paid by customer for purchasing.

So in Hadoop map reduce coding is done to get all customer id who had transaction more than 160. All the data we will get by specific key which is unique so than identification can be easily done and again process for more computation done to get only specific data is required coding .

Use case2: Getting Customer information in specific range

In the second use case, we are finding all the transaction where amount is between the range of170 and 200 according to customer ID, so that we come to know about the transactions in specific range , in specific area how much transaction happened.

Use case3:

In this use case finding sum of all the transaction amount and the counting of the transaction according to customer id.from this we can get the information about daily transaction and the information about all the customer in a day. Data can be used for annual analyses for a company profit.

Use case4: Calculate the average transaction value for each user id

In this case we will fetch the transaction amount data of each user and get the average of all transaction as per individual user id. Average rating for each user can be done periodically for analysis, customer choice analysis can be done by knowing number of product he/she purchased.

Use case 5: Division of single file into multiple files

In this use case dataset is divided into multiple sub file according to specific value. So hadop developers gets data from two table and divide them according to requirement .For example how many customer has used credit card for as a payment mode , How many customer took offer on products so that when season comes for discount we can inform those customer about discount offer.

Use case 6: the profession of user who has spent the maximum amount

In this use case we are given a task to find the name of profession from customer dataset to find maximum amount.

Usecase7: Finding Highest three spenders:

In this use case task is given to find only three highest spenders among all using two data set.

Use case 8: finding top three top customers:

In this use case we just want to fetch three customer report from datasets.

Usecase9: Special rewards

In this use case we are finding top three customers who has made highest transaction in particularly time period, rewards will be given to make like offer, coupon and attends special function free of cost.